Science Commons Tool Final Report

Version 0.1, APRIL 2007

Lead Investigator: Shan Chen

Prepared by:

Ron Chernich

Executive Summary

This is a report for the CR3 – "Science Commons Tool". The work package is part of the DART (Dataset Acquisition, Accessibility and Annotation e-Research Technologies) project.

The CR3 work package aims to reduce barriers to content acquisition by providing more rights options for science researchers. It applies Science Commons (SC) work to scientific research data and results, focusing on the Publishing area. It develops software that enables the attachment of standardized Creative Commons (CC) licenses. The investigated methods, incentives and technologies to motivate researchers to submit their research publications into institutional repositories.

As part of the CR3 work package, existing digital library repositories were examined, including Fedora, ePrintsUQ and Fez. As a result, a research prototype has been developed to demonstrate how Science Commons works, in the aspect of Publishing, namely SC Publishing.

A software prototype was built to demonstrate how publisher copyright policies can be integrated into SC Publishing. This is because we believe that researchers need to understand the rights of their publications before they issue a license for their work. If the publisher allows the publication to be archived, then the author/license creator can choose a Creative Commons (CC) license to grant more or fewer permissions for others to use their work under certain conditions, and upload the publication to the repository for others to access. By understanding their rights of publications, with a software tool to enable attachment of a license, researchers will have fewer barriers to content acquisition. The prototype demonstrates selection and attachment of CC licenses. It also allows the user to search and browse publications attached with different types of licenses in the repository.

Copyright checking is based on the SHERPA/RoMEO data. Hence it is not practical remote accessing RoMEO Web interface each time a user enters a publisher or a journal. To avoid internet traffic that might be caused by frequently remote querying the RoMEO Web site, the prototype populates the RoMEO data locally and performs periodic, remote, batch updates.

The prototype demonstrates a viable approach to creation and storage of SC/CC data from a desktop web interface. The service has known limitations. More research within the framework of an expanded user community is required to understand the impact of the limitations and improve the usability of the interface.

Table of Contents

1 Introduction

The purpose of this document is to outline the objectives of the work package CR3, the work that was performed and the outcomes and conclusions from the work.

The aim of CR3 is to help researchers select an appropriate Creative Commons (CC) license to facilitate sharing of their published work. This is explored through a prototype service built as a Java Language ™ Web Service for portability. Where possible, open source components are adapted to provide high level abstractions of the required subsystems. The architecture is illustrated in Fig 1.

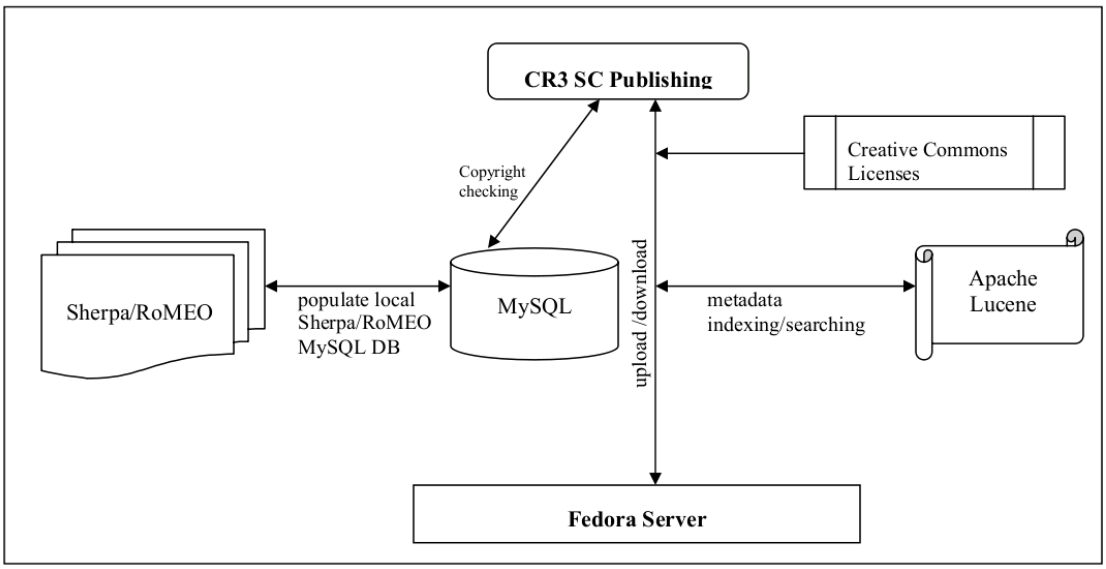

Fig 1 – CR3 Architecture

The objectives of the work package are to:

- Provide base level authentication
- Permit authenticated users to deposit publications by publisher or journal
- Associate an appropriate CC license with deposits.
- Maintain searchable metadata related to deposited papers
- Provide a search capability over deposited works by any combination of:
    - Creative Commons License
    - Publication Type
    - Year
    - Author

    - Title
    - Journal
    - Publisher
    - Abstract content
- Facilitate navigation from search results to specified URLs.

This document will describe the project's:

- Milestones,
- Outcomes including issues encountered,
- Technical Requirements and
- Recommendations.

2 Project Milestones

The initial milestones forCR3 included:

- Literature Review
- Requirements Analysis
- Development of prototype
- Demonstrate prototype to acquire feedback from representative end users
- Produce “video” demonstration package
- Document project outcomes and recommendations.

The development of this work package was developed by Shan Chen.

3 Project Outcomes

The outcomes of the project include:

- Java Web Service package (WAR file) that provides security, upload facility with associated metadata and search capability.
- Database management package
- Video demonstration of package in typical use case scenarios
- Documentation
- Interface to University of Queensland Fez database

3.1.1 Summary of Work

The current software being developed uses SHERPA/RoMEO as the resource for publisher copyright policies. To ensure researchers understand the rights of their publications, the software checks the publisher's conditions and/or restrictions before a publication is uploaded to the repository. If the publisher allows the publication to be archived, then the author/license creator can choose a CC license to grant more or fewer permissions for others to use their work under certain conditions. The software uses Fedora as the backend repository. It is allows searching of available publications under certain license(s).

3.1.2 Issues Encountered

Copyright checking is based on the remote RoMEO data. The prototype demonstrated the practical limitations of remote access to the RoMEO Web interface in real time each time a user enters a publisher or a journal. To avoid traffic and delays inherent in frequent remote querying the RoMEO Web site, it was necessary to mirror this data locally. A schedulable background task performs periodic synchronization of the local RoMEO data from the actual server.

3.2 Technical Requirements

The system comprises of the following components which all have various different technical requirements. These components include:

- Apache Tomcat
- Fedora over MySQL
- Sherpa/RoMEO API
- Apache Lucene text indexing and searching package

3.2.1 Apache Tomcat

Tomcat provide an open source Java Web Service “container” environment within which the CR3 service is deployed. Optionally, the service may be “front-ended” by a standard web server so that a static page can be accessed through the user’s web browser that redirects to the Tomcat server and port transparently.

3.2.2 Fedora

The Flexible Extensible Object Repository Architecture (Fedora) was developed as an open source project by the Cornell University and the University of Virginia Library. It provides a highly scalable object repository for the persisting of uploaded objects in any format and their retrieval. Full traceability with optional versioning are also provided. Fedora acts as a transaction broker for upload and indexed retrieval. It sits over a physical storage mechanism, generally a relational database management service (RDBMS). The selection of this service provides scalable performance and reliability.

3.2.3 Sherpa/RoMEO API

Sherpa provides an open access institutional repository. Their Rights Metadata for Open Archiving (RoMEO) project provides an API provides a machine to machine interface that facilitates storage and retrieval of permissions granted as part of a publisher's copyright agreement. This was initially intended to provide metadata harvested through the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) to reference centralized rights policies. The machine to machine interface provides a convenient way of providing reliable, centralized policies to be stored and retrieved.

3.2.4 Apache Lucene

The open source Lucene package provides a mechanism for full text indexing and searching. It provides a high level abstraction for swift searching of uploaded abstracts and other data.

4 Archival Storage of Project Deliverables

The CR3 source code is available via SVN (UQ internal only) at:
`svn+ssh://@datura/home/23/maenad/activities/dart/CR3/svn/CR3`
`svn+ssh://@datura/home/23/maenad/activities/dart/CR3/svn/RoMEO`
and can be browsed through a TRAC interface at:
http://datura/trac-cgi/trac.cgi/CR3/browser

A fully animated demonstration of all the service capabilities is available at:
http://www.itee.uq.edu.au/~eresearch/projects/dart/outcomes/cr3demo/cr3.html

All deliverable artefacts are stored on optical media for archival purposes.

5 Recommendations

The prototype service was constructed with the following constraints:

1. Author input is limited to three authors.
2. Proxy is set on the UQ ITEE based.
3. The RoMEO API for all publishers requires a full path specified in java code.
4. User password encryption is implemented on MySQL level.
5. Web client is tested on FireFox browser only.

These limitations are acceptable in a prototype where the installer can be expected to have the high level of technical understanding and skill required to make the changes needed to execute within their specific environment. A general, configurable end-user service would require refactoring and re-engineering to make these restrictions more easily configurable.

Potential future work includes:

- Dynamically generate multiple authors' input based on the user's requirement.
- Create configuration file for proxy setting and path to the RoMEO API local version (see 2. & 3. above).
- Implement comprehensive login session.
- Platform-independent password encryption.
- Allow paging on the search results.
- Integrate Ajax techniques for author/journal/publisher suggestion.
- Adjust the system to be used on different browsers.

6 Publications

(none)

7 Terms of Reference

7.1 Glossary

Acronym	Definition
API	Application Programming Interface
CC	Creative Commons
Fedora	Flexible Extensible Digital Object Repository Architecture
Fez	UQ front end to Fedora
JSF	Java Server Faces
OAI	Open Archives Initiative
PMH	Protocol for Metadata Harvesting
RDBMS	Relational Database Management System
RoMEO	Rights Metadata for Open Archiving
SC	Science Commons
UQ	University of Queensland
WAR	Web Archive

8 Report Signoff

It is agreed between

Lead Investigator

and

Chief Investigator

and

DART Project Director

That the **Final Report Document** for the CR3 – Scientific Commons gives a full account of the work undertaken for the DART Project.

Staff Member	Contact
Staff Member	Contact
Staff Member	Contact
Staff Member	Contact

- has been read and reviewed by all parties,
- shows that the work package CR3 has been completed satisfactorily,
- clearly outlines the functionality that was delivered.

Dated this ddth day of mmmm 20yy

Signed by name of CI for and on behalf of the Chief Investigator	Signed for and on behalf of DART by the Project Director Andrew Treloar